

## IMPLEMENTASI ALGORITMA K-MEANS PADA PERISTIWA GEMPA BUMI DI WILAYAH JAWA BARAT

Mirna Agustina<sup>1</sup>, Mulyawan<sup>2</sup>

[mirnaagustina243@gmail.com](mailto:mirnaagustina243@gmail.com)<sup>1</sup>, [wm7488748@gmail.com](mailto:wm7488748@gmail.com)<sup>2</sup>

<sup>1 2</sup> STMIK IKMI Cirebon

---

### Abstract

*Indonesia's strategic geological location causes frequent earthquakes, volcanic eruptions, and various natural disasters. West Java is a province that has quite a lot of natural disasters. From the collected data, natural disasters that often occur in West Java consist of several categories, including earthquakes. Many of these natural disasters have caused casualties, both fatalities and injuries, destroyed the surrounding area and its infrastructure, and caused property losses. The increase in the incidence of earthquakes needs to be further investigated to prevent the number of victims from increasing. This information can be obtained through a data mining approach, considering the large amount of data available. In relation to earthquake disaster data, clustering techniques in data mining are very useful for grouping disaster data based on the year of occurrence so that it can be used as a basis for predicting future earthquake events. Thus, this research produced 2 clusters, namely cluster 0 with 180 data in 2017 and a high incidence rate, and cluster 1 with 90 data and a low level in 2012. In the end, it is expected to be possible to classify earthquake disaster data using knowledge discovery data techniques (KDD) using the k-means algorithm seen at the time of the disaster.*

*Keywords: data mining; earthquake; natural disasters; K-means; knowledge discovery data.*

### Abstrak

Letak geologis Indonesia yang strategis menyebabkan sering dilanda gempa bumi, letusan gunung berapi dan berbagai bencana alam. Jawa Barat merupakan provinsi yang memiliki kejadian bencana alam cukup banyak, dari data yang dikumpulkan bencana alam yang sering terjadi di Jawa Barat terdiri dari beberapa kategori termasuk gempa bumi,. Banyaknya peristiwa bencana alam ini menyebabkan timbulnya korban, baik korban jiwa maupun luka-luka, menghancurkan wilayah sekitarnya serta menghancurkan infrastruktur dan menimbulkan kerugian harta benda. Adanya peningkatan kejadian bencana gempa bumi perlu diteliti lebih lanjut untuk mencegah agar jumlah korban tidak semakin banyak. Informasi tersebut dapat diperoleh melalui pendekatan data mining mengingat jumlah data yang tersedia sangat banyak. Dalam kaitannya dengan data bencana gempa bumi, teknik clustering pada data mining sangat berguna untuk mengelompokkan data bencana gempa bumi berdasarkan tahun kejadian sehingga dapat dijadikan dasar untuk memprediksi

peristiwa gempa bumi kedepannya. Dengan demikian, penelitian ini dihasilkan 2 cluster, yaitu cluster 0 sebanyak 180 data di tahun 2017 dengan tingkat kejadian tinggi dan cluster 1 sebanyak 90 data dengan tingkatan rendah pada tahun 2012. Pada akhirnya diharapkan dapat mengelompokkan data bencana gempa bumi memanfaatkan teknik knowledge discovery data (KDD) menggunakan algoritma k-means dilihat dari waktu terjadinya bencana.

**Kata Kunci:** Bencana alam; Gempa Bumi; Data Mining; K-Means; Knowledge Discovery Data

## Pendahuluan

Bencana alam adalah suatu peristiwa alam yang akan berdampak besar bagi kelangsungan hidup manusia. Salah satu bencana alam yang terjadi di Indonesia adalah gempa bumi, gempa bumi merupakan peristiwa getaran atau guncangan di permukaan bumi. Gempa bumi dapat disebabkan oleh pergerakan kerak bumi atau lempeng bumi, gunung berapi, dan tumbukan meteorit. Gempa bumi juga disebabkan oleh getaran lapisan batuan retak yang mengandung energi. Menurut beberapa sumber, gempa bumi yang disebabkan oleh pergerakan kerak bumi merupakan salah satu penyebab gempa bumi paling sering terjadi. Getaran tanah bervariasi dengan intensitas dari yang sangat kecil sehingga sulit dirasakan, sampai ke getaran yang sangat keras. Jawa Barat merupakan provinsi yang cukup banyak terjadi gempa, berdasarkan data yang dikumpulkan bencana gempa bumi di Jawa Barat menimbulkan korban, baik korban jiwa maupun luka-luka, meluluhlantakkan kawasan sekitar dan infrastruktur serta menimbulkan kerusakan harta benda. (Artatia & Hakim, 2015)

Perkembangan teknologi dan informasi saat ini mengalami perkembangan yang sangat pesat. Teknologi yang canggih membuat setiap orang dapat mengakses dan mendapatkan informasi secara cepat dan akurat, tanpa mengenal batasan wilayah dan waktu. Hal ini menjadikan informasi itu sangat berharga dan dibutuhkan untuk pengambilan keputusan. Salah satu informasi yang dibutuhkan oleh warga saat ini adalah informasi mengenai kejadian bencana gempa bumi. Namun informasi mengenai bencana alam yang tersaji saat ini masih bersifat acak, sulit dipahami dan belum terbukti keabsahannya. Meskipun bencana alam merupakan kejadian yang tidak bisa dihindari, akan tetapi dampak bencana dapat dikurangi dengan mengenali penyebab dan mempelajari kejadian bencana yang telah terjadi dengan menganalisa data bencana yang ada. Perlu dilakukan prediksi daerah rawan bencana alam supaya dapat menjadi indikator penanggulangan bencana alam. Berbagai metode pengolahan data menjadi sebuah informasi yang mudah dipahami sudah banyak dilakukan. Data mining merupakan teknik yang umum dilakukan untuk pengolahan data bencana alam. Teknik data mining dapat

diterapkan untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Dari polanya data mining di kelompokkan ke dalam beberapa Teknik yaitu deskripsi, kemudian estimasi, prediksi, klasifikasi, clustering, dan asosiasi. Data mining menjadi sebuah inti dari proses KDD (knowledge discovery in database) dimana KDD merupakan sebuah proses terorganisir untuk menemukan pengetahuan dan pola-pola yang berlaku dalam sebuah data yang besar. Data mining perlu dipelajari dan dipahami, karena manusia menghasilkan banyak data yang besar dalam berbagai bidang. (Joko Suntoro, 2019)

Pendekatan data mining bukanlah hal baru, terbukti dengan banyaknya penelitian terdahulu yang memanfaatkan. Beberapa penelitian terdahulu diantaranya, penelitian yang dilakukan oleh Muhamad Iqbal Ramadan dan Prihandoko dalam jurnal yang berjudul “Penerapan Data Mining Untuk Analisa Data Bencana Milik BNPB Menggunakan Algoritma K-means dan Linear Regression” menggunakan algoritma k-means untuk pengelompokan data dan linear regression untuk melakukan prediksi data bencana milik BNPB selama 5 tahun kedepan (Iqbal Ramadhan, 2017). Dalam jurnal berjudul “Implementasi K-means Clustering Untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model Crisp-dm” yang disusun oleh F N Dheawiyanti dkk, mengelompokkan daerah rawan bencana kebakaran dengan algoritma k-means clustering dengan model crispdm bertujuan menghasilkan kelompok daerah yang memiliki potensi tinggi terjadinya bencana kebakaran (Dhewayani et al., 2022). Dalam penelitian yang dilakukan Risa Halilintar dan Intan Nur Farida menggunakan algoritma k-means clustering untuk melakukan prediksi prestasi nilai akademik mahasiswa yang terbagi menjadi 4 kelompok atau cluster yaitu sangat baik, baik, cukup, kurang yang berguna untuk mengetahui mahasiswa yang mempunyai predikat kelulusan yang sesuai dengan pengelompokan tersebut.

Dataset yang digunakan untuk penelitian ini adalah data gempa bumi di Jawa Barat pada tahun 2012 hingga tahun 2021 sebanyak 270 data. Pada data tersebut terdapat tujuh atribut, dataset terkait Gempa Bumi ini dihasilkan oleh Badan Penanggulangan Bencana Daerah yang dikeluarkan dalam periode 1 tahun sekali. (Jabar, 2022)

Adanya peningkatan kejadian gempa bumi perlu diteliti lebih lanjut untuk mencegah agar jumlah korban tidak semakin banyak. Informasi tersebut dapat diperoleh melalui pendekatan data mining mengingat jumlah data yang tersedia sangat banyak. Dalam kaitannya dengan data gempa bumi, teknik clustering pada data mining sangat berguna untuk mengelompokkan data gempa bumi berdasarkan karakteristik yang sama sehingga dapat dijadikan dasar untuk memprediksi peristiwa gempa bumi kedepannya. (Algoritma & Pulau, 2022)

Berdasarkan latar belakang masalah, pada penelitian ini menggunakan teknik data mining untuk menggali informasi data kejadian gempa bumi dengan menggunakan algoritma K-Means Clustering untuk mengelompokkan kejadian gempa bumi yang ada di 27 Kabupaten/Kota di Provinsi Jawa Barat menjadi 2

cluster yaitu cluster rendah dan tinggi. Kemudian akan dilakukan pengelompokan kejadian bencana gempa bumi berdasarkan tahun di wilayah

h dengan intensitas kerawanan diatas 50% pada tiap variable bencana gempa bumi.

## Metode Penelitian

### Proses KDD (Knowledge Discovery Distance)

Tahapan perancangan mencakup langkah – langkah perancangan dari awal sampai akhir *Knowledge Discovery In Databases*. *Knowledge Discovery in Database Process* (KDD) adalah salah satu metode yang bisa digunakan dalam melakukan data mining. Fayyed et al. (1996) mendefinisikan KDD sebagai proses dari menggunakan metode data mining untuk mencari informasi-informasi yang berharga, pola yang ada di dalam data, yang melibatkan algoritma untuk mengidentifikasi pola pada data. Dunham (2003) meringkas proses KDD dari berbagai step, yaitu: seleksi data, pra-proses data, transformasi data, data mining, dan yang terakhir interpretasi dan evaluasi.

#### 1. Seleksi Data (Data Selection)

Pemilihan (seleksi) data dari sekumpulan data dengan mengidentifikasi semua sumber internal dan eksternal, kemudian hanya memilih sebagian saja dari data yang diperlukan dan disimpan suatu berkas terpisah dari data operasional untuk aplikasi data mining.

Seleksi ini dilakukan untuk mendapatkan kelompok data berdasarkan atribut yang diperlukan seperti Nama Kabupaten Kota, Kode Kabupaten Kota, Jumlah Kejadian Gempa, dan Tahun Kejadian.

#### 2. Pembersihan Data (Data Cleaning)

Pada proses ini, data-data yang tidak relevan dan tidak dibutuhkan akan dibersihkan atau dibuang agar data dapat digunakan pada proses berikutnya. Pada perancangan tugas akhir ini terjadi tahapan *cleaning* data, dimana atribut kode provinsi dan nama provinsi dihapus kemudian menjadi data untuk transformasi ke tahapan selanjutnya.

#### 3. Transformasi Data (Data Integration)

Integrasi data adalah penghubungan kembali data yang terpecah. Proses ini dilakukan dalam perancangan tugas akhir ini dikarenakan data yang disiapkan beberapa periode data.

#### 4. Transformasi Data

Pada tahap ini, kualitas dan hasil dan hasil data mining akan ditentukan. Atribut hasil dari *cleaning* dan mendapatkan proses reduksi akan digunakan sebagai dataset untuk pengujian perancangan tugas akhir ini.

#### 5. Data Mining

Pada proses ini data akan diproses dan menghasilkan *pattern*.

#### 6. Evaluasi

Pada tahap ini, proses KDD dilakukan, data mining yang dihasilkan adalah alur informasi yang ditampilkan dengan tampilan sederhana sehingga mudah dipahami.

### Penerapan K-Means Clustering menggunakan Rapidminer

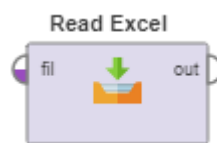
Pada tahapan ini meliputi pengumpulan data kejadian gempa bumi kemudian menerapkan algoritma *K-Means clustering* untuk mengelompokkan jumlah kejadian gempa bumi berdasarkan tahun. Dalam penerapan ini, menggunakan perhitungan *Euclidan Distance* dimana akan menghasilkan selisih jarak antara data satu dengan titik uji *centroid*. Jarak ini digunakan sebagai dasar penentuan pengelompokkan kejadian gempa bumi.

### Hasil dan Pembahasan

#### 1. Penerapan Algoritma K-Means Clustering Menggunakan Microsoft Excel

Pada perhitungan jarak ini dimuat sebanyak 2 iterasi, dengan nilai centroid awal yang diperoleh dari hasil proses Rapidminer dan *centroid* berikutnya diperoleh dari hasil iterasi 1. Dari perhitungan ini menghasilkan 2 Cluster, yaitu cluster 0 sebagai cluster tinggi sebanyak 180 data dan cluster 1 sebagai cluster rendah sebanyak 90 data.

#### 2. Proses dan Hasil Rapidminer



Gambar 1 Read Excel

Pada tahap ini adalah proses membaca file excel data kejadian gempa bumi yang telah di transformasi.



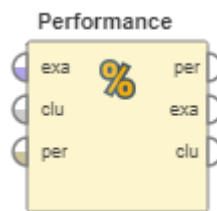
Gambar 2 K-Means Clustering

Clustering K-Means, tahap ini berfungsi untuk memperkirakan jarak antara centroid dengan 2 cluster.



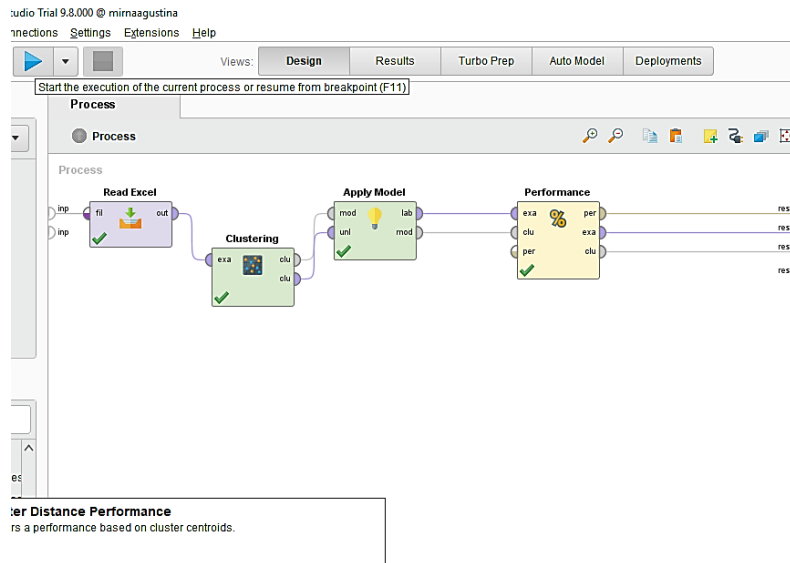
Gambar 3 Apply Model

Apply model, tahap ini untuk mendapatkan prediksi pada data yang tidak terlihat atau untuk mengubah data dengan menerapkan model preprocessing.



Gambar 4 Performance

Pada tahap ini dipilih operator cluster distance performance, Operator ini memberikan daftar nilai kriteria kinerja berdasarkan centroid cluster. mengambil model cluster centroid ini dan clustered set sebagai masukan dan mengevaluasi kinerja model berdasarkan centroid cluster untuk mendapatkan indeks Davies-Bouldin.



Gambar 5 Proses K- Means Clustering pada Rapidminer

## Kesimpulan

Berdasarkan dari pembahasan tugas akhir ini dapat disimpulkan hasil dari proses Rapidminer untuk implementasi algoritma *k-means clustering* pada data peristiwa gempa bumi di wilayah Jawa Barat, bahwa metode ini mampu menyelesaikan pengelompokkan *cluster* kejadian gempa bumi. Dengan hasil Rapidminer pula dapat diuji kembali secara manual di Microsoft Excel dengan menghitung jarak dan centroid dari *cluster* 0 dan *cluster* 1. Metode *K-Means Clustering* yang diterapkan pada rapidminer memiliki hasil yang sama dengan perhitungan manual menggunakan *Microsoft Excel* dengan jumlah C0 sebanyak 180 data (*cluster* tinggi) yang terjadi pada tahun 2017 dan C1 sebanyak 90 data (*cluster* rendah) yang terjadi pada tahun 2012 dengan hasil Davies Bouldin 0.058.

## Daftar Pustaka

- Algoritma, M., & Pulau, K. (2022). Klasterisasi Dampak Bencana Gempa Bumi Menggunakan Algoritma K-Means di Pulau Jawa. *Jurnal Edukasi Dan Penelitian Informatika*, 8(1), 175–179.
- Artatia, H., & Hakim, R. F. (2015). Pengelompokan Dampak Gempa Bumi dari Segi Kerusakan Fasilitas pada Provinsi yang Berpotensi Gempa di Indonesia Menggunakan K-Means-Clustering. *Prosiding Seminar Nasional Matematika Dan Pendidikan Matematika UMS*, 742–748.
- Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., & Jajuli, M. (2022). Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. *Jurnal Teknologi Dan Informasi*, 12(1), 64–77. <https://doi.org/10.34010/jati.v12i1.6674>

Iqbal Ramadhan, M. (2017). Penerapan Data Mining Untuk Analisis Data Bencana Milik Bnpb Menggunakan Algoritma K-Means Dan Linear Regression. *Jurnal Informatika Dan Komputer*, 22(1), 57–65.

Jabar, B. (2022). *sekilas-bpbd*. [Https://Bpbd.Jabarprov.Go.Id/](https://Bpbd.Jabarprov.Go.Id/).  
<https://bpbd.jabarprov.go.id/sekilas-bpbd/>

Joko Suntoro. (2019). *Data Mining: Algoritma dan Implementasi dengan Pemrograman PHP* (pertama). PT Elex Media Komputindo.